Aggregate Measures: Development, Creation, and Validation

Indiana University Bloomington, Center for Postsecondary Research

Allison BrckaLorenz, Kyle T. Fassett, Thomas Kirnbauer

Overview

Quantitative and survey research relies heavily on sound design to be able to understand student as well as faculty experiences. Often IR professionals and survey researchers are tasked with investigating broad concepts such as climate or engagement. Determining where to start on such an endeavor can be daunting. This poster aims to provide tips and suggestions for operationalizing broad concepts, writing survey items, creating aggregate measures (scales, indexes, etc.), and evaluating the quality and validity of those measures.

About FSSE

The Faculty Survey of Student Engagement (FSSE) was implemented in 2003 to measure instructional staff expectations for student engagement in educational practices that are empirically linked with high levels of learning and development. Student engagement is the intentional time and effort dedicated to improving student learning in the classroom.

Framework for Validation

Validation evidence for aggregate measures can come in many forms:

Survey content: Evidence showing FSSE's content is representative of specific aspects of faculty practice

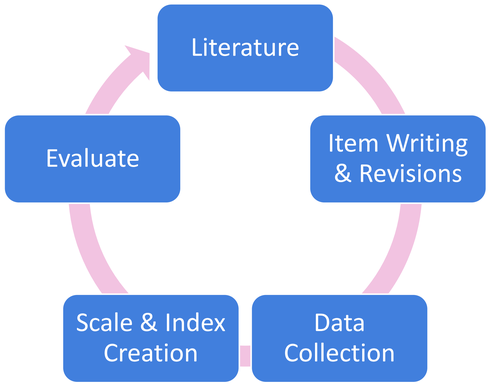
Response process: Evidence indicated that respondents interpret and respond to FSSE as intended

Internal structure: Evidence indicating the degree to which questions relate to dimensions of faculty practice

Relations to other measures: Evidence indicating the degree to which relationships with other measures follow expectations

Consequences of assessment: Evidence indicating the degree to which FSSE data use results in expected outcomes

Development Process



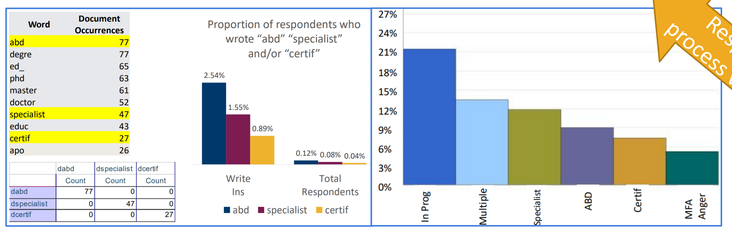
Item Writing & Revisions

Survey content validity

Quality aggregate measures require quality component items:

- Solidify the concept you want to study. Scour the literature and expert opinions for specific behaviors, perceptions, and attitudes that typify your concept. Create an exhaustive list before editing down for redundancy. Remove any ideas that will not contribute to your goals.
- Decide the perspective with which you wish to investigate your concept to create your question stem: self-reported participation in behaviors or activities, agreement with pre-determined statements, degree of importance or value, extent of perception, etc.
- Adhere to best practices for item writing and survey construction to refine your items .
- Conduct cognitive interviews or focus groups to ensure respondents understand your items the way you intend. Have experts, non-experts, and members of the group of people you intend to study review the items for clarity. Pilot test items if possible and analyze any write-in information to look for signs of misunderstanding.

Response process validity



(Yuhas & BrckaLorenz, 2018)

Scale & Index Creation

Exploratory factor analysis, confirmatory factor analysis, and Cronbach's α are common statistical methods used to evaluate the quality and validity of combining a group of items into an aggregate measure. Other indicators such as item means and standard deviations, inter-item correlations, and intraclass correlation coefficients (ICC) can also be helpful indicators in creating a scale. When calculating scale scores, consider what range you would like your scale to have, either the original range of individual items or transformed into a different rage. FSSE items, for example have single digit ranges but FSSE scales are transformed to a 0-60 range. Be cautious about transforming scale ranges to a 0-100 range as people will likely misinterpret scores as percentages! If you have binary (yes/no, done/not done, etc.) items, considering combining them into an additive index where scores represent the number of positive responses or participation in certain experiences.

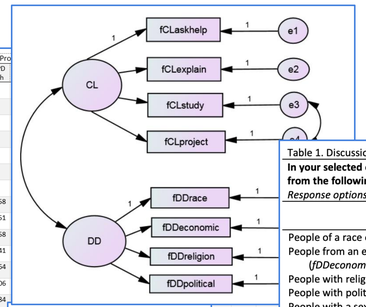
Internal structure validity

Table 4. Significant Correlations ($p < .01$) between Teaching Pro

	fTPD 01a	fTPD 01b	fTPD 01c	fTPD 01d	fTPD 01e	fTPD 01f	fTPD 01g	fTPD 01h															
1b	.688																						
1c	.631	.624																					
1d	.525	.496	.593																				
1e	.475	.351	.438	.403																			
1f	.520	.518	.515	.500	.499																		
1g	.611	.562	.597	.472	.477	.544																	
1h	.554	.480	.590	.442	.404	.400	.643																
1i	.510	.445	.576	.443	.371	.385	.554	.758															
1j	.440	.420	.526	.438	.388	.458	.474	.551															
1k	.446	.481	.498	.447	.400	.494	.497	.458															
1l	.537	.462	.575	.456	.405	.405	.635	.741															
1m	.519	.417	.554	.398	.371	.331	.564	.764															
1n	.524	.431	.583	.446	.355	.352	.534	.706															
3a	.229	.152	.193	.178	.180	.153	.169	.184															
3b	.271	.175	.199	.218	.209	.183	.209	.211															
3c	.149	.121	.137	.155	.103	.091	.130	.179	.185	.141	.130	.176	.174	.186	.267	.303							
3d	.209	.149	.204	.213	.166	.148	.199	.237	.229	.187	.156	.228	.236	.247	.309	.325	.499						
3e	.224	.153	.205	.199	.166	.145	.192	.248	.240	.180	.151	.229	.252	.260	.348	.449	.413	.614					
3f	.080	.082	.042*	.095	.056	.111	.034*			.063	.073				.173	.299	.235	.309	.312				
3g	.252	.162	.184	.220	.219	.195	.184	.149	.156	.133	.184	.174	.166	.182	.254	.373	.218	.267	.30				
3h	.213	.157	.193	.193	.196	.186	.173	.159	.157	.186	.175	.162	.159	.175	.213	.266	.256	.263	.246	.320	.442		

Note: Correlations are significant ($p < .01$) unless indicated $^{*}p < .05$.

(Strickland & BrckaLorenz, 2019)



(Paulsen & BrckaLorenz, 2019a)

Table 1. Discussions with Diverse Others Item Descriptives

In your selected course section, how much opportunity do students have to engage in discussions with people from the following groups?

Response options: 4=*Very much*, 3=*Quite a bit*, 2=*Some*, 1=*Very little*

	Count	Mean	Std. Dev.	Factor Loading
People of a race or ethnicity other than their own (*fDDrace*)	42,961	2.57	1.078	.927
People from an economic background other than their own (*fDDeconomic*)	42,596	2.56	1.003	.923
People with religious beliefs other than their own (*fDDreligion*)	42,289	2.38	1.029	.907
People with political views other than their own (*fDDpolitical*)	42,137	2.45	0.988	.898
People with a sexual orientation other than their own (*fddsexorient*)	41,748	2.19	1.016	--

Table 4. Discussions with Diverse Others Scale Descriptives

Scale	Count	Minimum	Maximum	Mean	Std. Dev.	Cronbach's Alpha	ICC
fDD	42292	0	60	29.77	18.729	.934	.132

(Priddie & BrckaLorenz, 2019)

Evaluate

Further analysis using your aggregated measure can provide further validation of its quality. This can be done by examining correlations of your measure with other measures of constructs known to be related from the literature. Looking for known-differences by subgroups of your population can also provide evidence of validation. Differential Item Functioning, for example, is a method that can be used to determine whether or not subgroups of individuals are answering questions in similar ways. Many other forms of evaluation can be specific to concerns you may have about your data collection such as for social desirability and nonresponse bias.

Relations to other variables

Table 5. Significant ($p < .001$) Correlations between Discussions with Diverse Others Scale and F

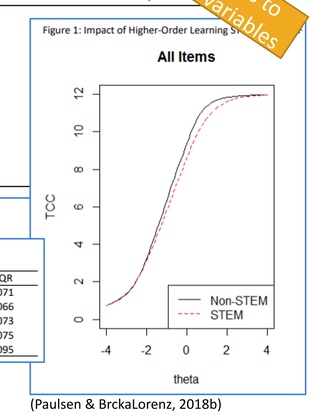
	fDD
Higher-Order Learning	.285
Reflective & Integrative Learning	.314
Learning Strategies	.189
Quantitative Reasoning	.077
Collaborative Learning	.156
Student-Faculty Interaction	.167
Effective Teaching Practices	.235
Quality of Interactions	.058
Supportive Environment	.189

(Priddie & BrckaLorenz, 2019)

Table 1. Pearson's Product Moment Correlations between FSSE and External Scales

		FSSE Scales					
		fET	fRI	fCL	fDD	fHO	fQR
External Scales	CCI		.479			.451	.071
	SPS		.420	.235			.066
	CB					.465	.073
	AL	.410			.317		.075
	ORG	.514					.095

(Paulsen & BrckaLorenz, 2018c)



(Paulsen & BrckaLorenz, 2018b)

Knowing that data and results from your measure are being used (and useful!) in ways can provide validation evidence as well.

Consequences of assessment

- FSSE's Publications and Presentations webpage displays research using FSSE data (http://fsse.indiana.edu/html/publications_presentations.cfm)
- FSSE's Psychometric Portfolio contains additional studies providing evidence for the validation of FSSE data and results (http://fsse.indiana.edu/html/Psychometric_Portfolio.cfm)
- Our *Lessons from the Field* publications and searchable database contain stories of ways that participating institutions are using FSSE data to improve the undergraduate experience (http://nsse.indiana.edu/html/howInstitutionsUseNSSE.cfm)

References & Resources

- Fassett, K. T. & BrckaLorenz, A. (2019). Scholarship of teaching and learning. FSSE Psychometric Portfolio. Retrieved from fsse.indiana.edu.
- Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 25* (2), 186-192.
- Paulsen, J. & BrckaLorenz, A. (2018a). Construct validity: 2017 scales. *FSSE Psychometric Portfolio*. Retrieved from fsse.indiana.edu.
- Paulsen, J. & BrckaLorenz, A. (2018b). Known-groups validity: 2017 FSSE measurement invariance. *FSSE Psychometric Portfolio*. Retrieved from fsse.indiana.edu.
- Paulsen, J. & BrckaLorenz, A. (2018c). Relations to other variables: 2018 convergent & divergent correlations. *FSSE Psychometric Portfolio*. Retrieved from fsse.indiana.edu.
- Priddie, C. & BrckaLorenz, A. (2019). Discussions with Diverse Others. *FSSE Psychometric Portfolio*. Retrieved from fsse.indiana.edu.
- Strickland, J. & BrckaLorenz, A. (2019). Teaching Professional Development. *FSSE Psychometric Portfolio*. Retrieved from fsse.indiana.edu.
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806-838.
- Yuhas, B. & BrckaLorenz, A. (2018). Response process validity: 2013-2016 write-in response analysis. *FSSE Psychometric Portfolio*. Retrieved from fsse.indiana.edu.

@NSSEsurvey